9491 IJTO-1 PCT/EP 94/01323

1

Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

93401099.2

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

H. van Bilderbeek

Den Haag, den
The Hague, 27/06/94
La Haye, le

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr
Application no
Demande n° 93401099.2

Anmeldetag
Date of filing
Date de dépôt 27/04/93

Name(n) der Anmelder(in)(e)
Name(s) of applicant(s)
Nom(s) du(des) demandeur(s)
N.V. INNOGENETICS S.A.
B-9052 Gent
BELGIUM

Bezeichnung der Erfindung
Title of the invention
Titre de l'invention Sequences of hepatitis C virus genotypes and their use as therapeutic and diagnostic agents

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat State Pays	Tag Date Date	Aktenzeichen File no Numéro de dépôt

Internationale Patentklassifikation
International Patent classification
Classification internationale des brevets
C12N15/51, C12Q1/68, C12Q1/70, A61K39/29, G01N33/576, C07K7/04

An Anmeldetag benannte Vertragstaaten
Contracting states designated at date of filing
Etats contractants désignés lors du dépôt AT/BE/CH/DE/DK/ES/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE

Bemerkungen
Remarks
Remarques The title of the application as originally filed reads as follows:
"New sequences of hepatitis C virus genotypes and their use as therapeutic and diagnostic agents".

2

NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS THERAPEUTIC AND DIAGNOSTIC AGENTS

The invention relates to new sequences of hepatitis C virus genotypes and their use as therapeutic and diagnostic agents.

The present invention relates to new nucleotide and amino acid sequences corresponding to type-specific regions of Hepatitis C virus type 3 and the coding region of Hepatitis C virus type 4, a process for preparing them, and their use for diagnosis, prophylaxis and therapy.

The technical problem underlying the present invention is to provide new type-specific sequences of the Core, the E1, the NS3, the NS4 and the NS5 regions of HCV type 3 and type 4. These new HCV sequences are useful to diagnose the presence of type 3 and/or type 4 HCV genotypes present in a biological sample. Moreover, the availability of these new type-specific sequences can increase the overall sensitivity of HCV detection and should also prove to be useful for therapeutic purposes.

Hepatitis C viruses (HCV) have been found to be the major cause of non-A, non-B hepatitis. The sequences of cDNA clones covering the complete genome of several prototype isolates have already been determined (Kato et al., 1990; Choo et al., 1991; Okamoto et al., 1991; Okamoto et al., 1992). Comparison of these isolates shows that the variability in nucleotide sequences can be used to distinguish at least 2 different genotypes, type 1 (HCV-1 and HCV-J) and type 2 (HC-J6 and HC-J8), with an average homology of about 68%. Within each type, at least two subtypes exist (e.g. represented by HCV-1 and HCV-J), having an average homology of about 79%. HCV genomes belonging to the same subtype show average homologies of more than 90% (Okamoto et al., 1992). However, the partial nucleotide sequence of the NS5 region of the HCV-T isolates showed at most 67% homology with the previously published sequences, indicating the existence of a new type (Mori et al., 1992). Parts of the 5' untranslated region (UR), core, NS3, and NS5 regions of this type 3 have been published, further establishing the similar evolutionary distances between the 3 major genotypes and their subtypes (Chan et al., 1992).

The identification of type 3 genotypes in clinical samples can be achieved by means of PCR with type-specific primers for the NS5 region. However, the degree to which this will be successful is largely dependent on sequence variability and on the virus titer present in the serum. Therefore,

The LiPA format is completely compatible with commercially available scanning devices, thus rendering automatic interpretation of the results very reliable. All those advantages make the LiPA format liable for the use of HCV detection in a routine setting. The LiPA format should be particularly advantageous for detecting the presence of different HCV genotypes.

The present invention also relates to a method for detecting and identifying novel HCV genotypes, different from the known HCV genomes, comprising the steps of:

- determining to which HCV genotype the nucleotides present in a biological sample belong, according to the process as defined above,
- in the case of observing a sample which does not generate a hybridization pattern compatible with those defined in Table 3, sequencing the portion of the HCV genome sequence corresponding to the aberrantly hybridizing probe of the new HCV genotype to be determined.

The present invention also relates to the use of a composition as defined above, for detecting one or more genotypes of HCV present in a biological sample liable to contain them, comprising the steps of:

(i) possibly extracting sample nucleic acid,

(ii) amplifying the nucleic acid with at least one of the primers as defined above,

(iii) sequencing the amplified products

(iv) inferring the HCV genotypes present from the determined sequences by comparison to all known HCV sequences.

The present invention also relates to a composition consisting of or comprising at least one peptide or polypeptide comprising a contiguous sequence of at least 5 amino acids corresponding to an amino acid sequence encoded by at least one of the HCV genomic regions as defined above, having at least one amino acid differing from the corresponding region of HCV type 1 and/or type 2 polyprotein sequences, or muteins thereof.

The new type 3 amino acid sequences, as deduced from the disclosed nucleotide sequences (see SEQ ID NO 1 to 42), show homologies of only 59.9 to 78% with prototype sequences of type 1 and 2 for the NS4 region, and of only 53.9 to 68.8% with prototype sequences of type 1 and 2 for the E1 region. As the NS4 region is known to contain several epitopes, for example characterized in patent application EP-A-0 489 968, and as the E1 protein is expected to be subject to immune attack as part of the viral envelope and expected to contain epitopes, the NS4 and E1 epitopes of the new type 3 and 4

vol. 15-I et II. THIEME, Stuttgart 1974.

The polypeptides of the invention can also be prepared in solid phase according to the methods described by Atherton and Shepard in their book entitled "Solid phase peptide synthesis" (IRL Press, Oxford, 1989).

The polypeptides according to this invention can be prepared by means of recombinant DNA techniques as described by Maniatis et al., Molecular Cloning; A Laboratory Manual, New York, Cold Spring Harbor Laboratory, 1982).

The present invention relates more particularly to a composition as defined above, with said polypeptide or peptide having at least one of the following amino acids in its peptidic chain:

- A157, F182, I186, H187, A190, S191 or G191, L192, W194, V202, L203, V219, I227, Q231, T237 or A237, T240, Y250, T254, S260, M271, M280, Q299, T303, L308, and L313 for the Core/E1 region, and D1556, Q1579, L1581, S1584, F1585, E1606, V1612, P1630, C1636, T1656, L1663, H1685, E1687, G1689, Y1705, A1714, A1721, V1723, H1726, R1738, Q1743, A1744, E1747, I1749, A1751, A1759 and H1762 for the NS3/NS4 region, as detected in type 3 sequences of the present invention,

- M44, Q70, A87, N106, K115, G142, I144, I178, P193, Y194, A197, M231, T232, V235, I242, S247, P249, S250, L251, V254, P257, A261, Y264, A266, G268, A280, L284, Y293, Q297, A299, and N303 in the Core/E1 region, and H1310, V1312, Q1321, P1368, V1372, N1399, F1648, P1651, V1667, T1669, A1681, A1700, Q1704, A1713, S1714, M1718, D1719, T1721, R1722, A1723, G1726, F1735, I1736, S1737, T1739, G1740, K1742, T1745, L1746, K1747, A1750, V1753, N1755, A1757, D1758, T1763, and Y1764 for the NS3/NS4 region, as detected in type 4 sequences of the present invention,

- D217, A213, A256, R294, V1677, Q1704, E1730, V1732, Q1741 and T1751 for the NS3/NS4 regions, as detected in type 3 and 4 sequences of the present invention, and with said notation being composed of a letter, unambiguously representing the amino acid by its one-letter code, and a number representing the amino acid numbering according to Kato et al., 1990 (see also Table 1 for comparison with other isolates).

For example M231 refers to a methionine at position 231. A glutamine (Q) is present at the same position 231 in type 3 isolates, whereas this position is occupied by an arginine in type 1 isolates and by a lysine (K) or asparagine (N) in type 2 isolates (see Figure 1A).

The peptide or polypeptide according to this embodiment of the

acids Q299 and T303 are unique for type 3 isolates. The type 4 isolate shows the following unique V5 sequence: RPRQHATVQN (SEQ ID NO 92), of which Q297, A299, and N303 are unique for type 4. Amino acid R294 is unique for type 3 and 4 isolates.

Consequently, the present invention also relates to a composition as defined above, wherein said peptides or polypeptides contain in their peptidic chain an amino acid sequence selected from any of the regions spanning the following positions of HCV type 3 polyproteins:

- positions 140 to 319 in the Core/E1 region, more particularly a composition wherein said polypeptide or peptide corresponds to a sequence within any of the amino acid sequences as represented in SEQ ID NO 14, 16, 18, 20, 22, 24, 26 or 28, or any other HCV amino acid sequence having a homology of more than 69%, preferably more than 70%, and most preferably more than 72% in the E1 region spanning positions 192 to 319 to any of the amino acid sequences as represented in SEQ ID NO 14, 16, 18, 20, 22, 24, 26 or 28, preferably a composition containing at least one of the following polypeptides:

LEWRNTSGLYVL (SEQ ID NO 83), VYEADDVILHA (SEQ ID NO 85), VQDGNTST (SEQ ID NO 94), VQDGNTSA (SEQ ID NO 95), VQDGNTSTCWTPV (SEQ ID NO 87), VKYVGATTAS (SEQ ID NO 96), VRYVGATTAS (SEQ ID NO 89), RPRRHQTVQT (SEQ ID NO 91), or any synthetic peptide or polypeptide containing at least 5 contiguous amino acids derived from the above-defined peptides in their peptidic chain.

- positions 1646 to 1764 in the NS3/NS4 region, more particularly a composition wherein said peptide or polypeptide corresponds to a sequence within any of the amino acid sequences as represented in SEQ ID NO 30, 32, 34, 36, 38 or 40, or any other HCV amino acid sequence having a homology of more than 76%, preferably more than 78%, most preferably more than 80% to any of the amino acid sequences as represented in SEQ ID NO 30, 32, 34, 36, 38 or 40, in the region spanning positions 1646 to 1764, preferably a composition containing at least one of the following polypeptides:

LGGKPAIVPDKEVLYQQYDE (SEQ ID NO 97),
LGGKPALVPDKEVLYQQYDE (SEQ ID NO 98),
SQAAPYIEQAQVIAHQFKEK (SEQ ID NO 99),
IAHQFKEKVLGLLQRATQQQ (SEQ ID NO 100),
IAHQFKEKILGLLQRATQQQ (SEQ ID NO 101),

or any synthetic peptide or polypeptide containing at least 5 contiguous

(iii) ANTI-SENSE: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: BR36-20-164

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..401

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

```
TC CAA AAT GAA ATC TGC TTG ACA CAC CCC ATC ACA AAA TAC ATC ATG          47
   Gln Asn Glu Ile Cys Leu Thr His Pro Ile Thr Lys Tyr Ile Met
     1               5                  10                  15

GCA TGC ATG TCA GCT GAT CTG GAA GTA ACC ACC AGC ACC TGG GTT TTG         95
Ala Cys Met Ser Ala Asp Leu Glu Val Thr Thr Ser Thr Trp Val Leu
                20                  25                  30

CTT GGA GGG GTC CTC GCG GCC CTA GCG GCC TAC TGC TTG TCA GTC GGT        143
Leu Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Val Gly
            35                  40                  45

TGT GTT GTG ATT GTG GGT CAT ATC GAG CTG GGG GGC AAG CCG GCA ATC        191
Cys Val Val Ile Val Gly His Ile Glu Leu Gly Gly Lys Pro Ala Ile
        50                  55                  60

GTT CCA GAC AAA GAG GTG TTG TAT CAA CAA TAC GAT GAG ATG GAA GAG        239
Val Pro Asp Lys Glu Val Leu Tyr Gln Gln Tyr Asp Glu Met Glu Glu
    65                  70                  75

TGC TCA CAA GCT GCC CCA TAT ATC GAA CAA GCT CAG GTA ATA GCT CAC        287
Cys Ser Gln Ala Ala Pro Tyr Ile Glu Gln Ala Gln Val Ile Ala His
80                  85                  90                  95

CAG TTC AAG GGA AAA GTC CTT GGA TTG CTG CAG CGA GCC ACC CAA CAA        335
Gln Phe Lys Gly Lys Val Leu Gly Leu Leu Gln Arg Ala Thr Gln Gln
                100                 105                 110

CAA GCT GTC ATT GAG CCC ATA GTA ACT ACC AAC TGG CAA AAG CTT GAG        383
Gln Ala Val Ile Glu Pro Ile Val Thr Thr Asn Trp Gln Lys Leu Glu
            115                 120                 125

GCC TTT TGG CAC AAG CAT                                                401
Ala Phe Trp His Lys His
        130
```

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

```
Gln Asn Glu Ile Cys Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala
```

```
 1               5                  10                  15

Cys Met Ser Ala Asp Leu Glu Val Thr Thr Ser Thr Trp Val Leu Leu
            20                  25                  30

Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Val Gly Cys
        35                  40                  45

Val Val Ile Val Gly His Ile Glu Leu Gly Gly Lys Pro Ala Ile Val
    50                  55                  60

Pro Asp Lys Glu Val Leu Tyr Gln Gln Tyr Asp Glu Met Glu Glu Cys
65                  70                  75                  80

Ser Gln Ala Ala Pro Tyr Ile Glu Gln Ala Gln Val Ile Ala His Gln
                85                  90                  95

Phe Lys Gly Lys Val Leu Gly Leu Leu Gln Arg Ala Thr Gln Gln Gln
            100                 105                 110

Ala Val Ile Glu Pro Ile Val Thr Thr Asn Trp Gln Lys Leu Glu Ala
        115                 120                 125

Phe Trp His Lys His
    130
```

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 401 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vii) IMMEDIATE SOURCE:
(B) CLONE: BR36-20-166

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..401

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

```
TC CAA AAT GAA ATC TGC TTG ACA CAC CCC ATC ACA AAA TAC ATC ATG          47
   Gln Asn Glu Ile Cys Leu Thr His Pro Ile Thr Lys Tyr Ile Met
     1               5                  10                  15

GCA TGC ATG TCA GCT GAT CTG GAA GTA ACC ACC AGC ACC TGG GTT TTG          95
Ala Cys Met Ser Ala Asp Leu Glu Val Thr Thr Ser Thr Trp Val Leu
                 20                  25                  30

CTT GGA GGG GTC CTC GCG GCC CTA GCG GCC TAC TGC TTG TCA GTC GGT         143
Leu Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Val Gly
             35                  40                  45
```

(B) MAP POSITION: positions 248 to 257 of the V4 region of HCV type 3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

Val Lys Tyr Val Gly Ala Thr Thr Ala Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: BR36

(viii) POSITION IN GENOME:
(B) MAP POSITION: Positions 1688 to 1707 of HCV type 3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

Leu Gly Gly Lys Pro Ala Ile Val Pro Asp Lys Glu Val Leu Tyr Gln
1 5 10 15

Gln Tyr Asp Glu
20

(2) INFORMATION FOR SEQ ID NO: 98;

(i) SEQUENCE CHARACTERISTICS;
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: HD10

(viii) POSITION IN GENOME:
(B) MAP POSITION: positions 1688 to 1707 of HCV type 3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

Leu Gly Gly Lys Pro Ala Leu Val Pro Asp Lys Glu Val Leu Tyr Gln
1 5 10 15

Gln Tyr Asp Glu
20

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(viii) POSITION IN GENOME:
(B) MAP POSITION: positions 1712 to 1731

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

Ser Gln Ala Ala Pro Tyr Ile Glu Gln Ala Gln Val Ile Ala His Gln
1 5 10 15

Phe Lys Glu Lys
20

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: BR36

(viii) POSITION IN GENOME:
(B) MAP POSITION: positions 1724 to 1743 of HCV type 3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

Ile Ala His Gln Phe Lys Glu Lys Val Leu Gly Leu Leu Gln Arg Ala
1 5 10 15

Thr Gln Gln Gln
20

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO